BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

R E M A R K S

Claims 1-11, 13-16 and 18 are pending and stand ready for further action on the merits.

The specification and claims have been amended for clarity.

Support for new claim 18 can be found in claim 12.

No new matter has been added by way of the above-amendment.

The above-amendment has not been made for the sake of patentability and/or does not narrow the scope of the invention.

[I] INTERVIEW

Applicants note with appreciation that the Examiner conducted an interview with Applicants' representative on February 4, 2004. The Examiner was very helpful in clarifying the outstanding issues.

On the Interview Summary Form, the Examiner states:

> The attorney was informed that a computer program "per se" is not statutory. Amendments to overcome this problem were discussed. Specific claim steps were reviewed with the intent to improve clarity and accurately reflect the disclosure of the specification and figures. No particular agreement for amendments was reached.

Further details of the interview are now provided.

[II] Issues under 35 U.S.C. 101 and 112, 1st paragraph

Claims 1-12 are rejected under 35 U.S.C. 101 as being drawn to non-statutory subject matter; and claims 1-12 are rejected under 35

U.S.C. 112, 1st paragraph, since the specification does not provide sufficient enablement for the skilled artisan to use the claimed invention.

Applicants respectfully traverse each of the rejections.

Applicants herein incorporate by reference the comments made regarding this issue in the January 7, 2004 Reply in their entirety.

During the Interview, the Examiner took the position that, in part, the lack of clarity of the claims made it difficult to find that the utility and enablement requirements have been met. The following comments in sections (II)(A)-(II)(C) address the Examiner's specific objections made during the Interview.

(II)(A) Sequential Assignment of Atoms to Hierarchical Levels:

During the interview, the Examiner took the position that the "classifying step" is not clear, since the step of "sequentially assigning atoms to hierarchical levels" possibly encompasses the creation of a second string of atoms starting with a new "first" atom. Applicants have amended the "classifying" step to clarify that a second string of atoms is not created starting with a new "first" atom in the classifying step.

However, it is important to note that a conformation code is prepared based on a molecular tree obtained with a single start atom. The present invention includes embodiments wherein additional conformation codes can be prepared using molecular trees formed using additional individual start atoms. Further details of this aspect of the invention is given below in Section (II)(B).

(II)(B) Number of Start Atoms:

During the interview, the Examiner expressed that she was unclear as to the number of start atoms which can be used in the method of the presently claimed invention.

For a molecule consisting of n atoms, there can be n starting atoms. By setting one atom as a starting atom, the procedures of classifying atoms, forming a molecular tree, coding the one atom and expressing the molecular tree listed in claim 1 are carried out.

Further, by setting another (a second) atom as another starting atom, the procedures of classifying atoms, forming a molecular tree, coding one atom and expressing said molecular tree are carried out. This is expressed as "and similarly, preparing conformation codes with respect to other start atoms" in the last clause of claim 1. These procedures can be fundamentally repeated for n starting atoms.

In some cases, it might be enough to set only special atoms as starting atoms. However, for example, in the case of predicting NMR shifts, it is necessary to set all n atoms as starting atoms and to repeat these procedures n times to be able to predict the NMR chemical shift of all nuclei. Its practical usage is described in Satoh et al. (J. Comput. Aided Chem., 2002), which was submitted with the January 7, 2004 Reply.

More precisely, the prediction of NMR chemical shifts is carried out by selecting the necessary codes from all the codes established in the n molecular trees, because some atoms are physically and electronically so distant as to not effect the NMR chemical shift. This is detailed in the aforementioned Satoh et al. (J. Comput. Aided Chem., 2002).

(II)(C) Description of Rules in Claims:

During the interview, the Examiner requested further clarification of the following rules which appear in the claims: a) "precedence rule"; b) "linear notation rule"; and c) "angle dividing rule."

(II)(C)(i) "Precedence Rule" and "Linear Notation Rule":

As noted during the interview, one embodiment of the present invention incorporates the CANOST code for the *"Precedence Rule" and "Linear Notation Rule."*

However, the Examiner noted that the inventive claims are drawn to a computer program for molecular stereochemical coding *any* molecule. The claims are not limited to a certain class of molecules. As such, the Examiner finds fault in the fact that the CANOST code list of Figure 4 does not include every conceivable atomic structure for every atom.

Applicants respectfully submit that the Examiner's position is not correct with respect to the CANOST code list of Figure 4. The Examiner's attention is directed to the "code X" in the CANOST code list of Figure 4, which can be used to represent any "other atom." Therefore, it can be said that any molecular structure can be expressed by using the CANOST code list of Figure 4 and that the CANOST code list of Figure 4 includes every conceivable atomic structure for every atom.

Furthermore, Applicants enclose herewith a journal article from Abe et al., (J. Chem. Inf. Comput. Sci., Vol. 24, No. 4, pg 212, **1984**) for the Examiner's review that describes the use of a CANOST code list. The only difference between the inventive CANOST code list of Figure 4, is the addition of "code H" to the inventive

list. In Abe et al., the procedure for forming a molecular tree is set forth.

With respect to the "Precedence Rule," the Examiner is respectfully requested to look to page 11, lines 21-28 wherein the precedence is defined in the column titled "no." of Figure 4. Accordingly, using the non-limiting embodiment of the CANOST precedence rule, the skilled artisan would assign the values given in the "no." column of Figure 4 for each of the substructures provided in the "substructure" column of Figure 4.

(II)(C)(ii) "Angle Dividing Rule"

With respect to the "angle dividing rule," Applicants have amended the specification at page 14, lines 3-4 to indicate that this rule is defined in Figures 6+ of the present application.

(II)(D) Requirement for "Tangible" Result:

As noted by the Examiner on the Interview Summary Form (reproduced above), the Examiner has taken the position that a computer program not claimed as embodied by computer-readable media is descriptive material *per se* and is not statutory because it is not capable of causing a functional change in the computer.

In response, Applicants have amended claim 1 to recite the subject matter of claim 16, i.e., that the conformation code of the molecule is printed on computer readable recording media.

Furthermore, the Examiner cited MPEP §2106(II)(A) which indicates as follows:

> [t]he claimed invention as a whole must accomplish a practical application. That is, it must produce a "useful, concrete and tangible result." (Citation omitted). The purpose of this requirement is to limit patent protection to inventions that possess a certain level of "real world" value, as opposed to subject matter that represents nothing more than an idea or concept, or is simply a starting point for future investigation or research. (Citations omitted).

The Examiner has taken the position that the inventive program just manipulates data and transforms the data into a different form, and as such, is merely a concept with no utility.

First, Applicants respectfully submit that MPEP 2106(II)(A) sets forth the requirement that the *specification* provides a complete disclosure which contains some indication of the practical application for the claimed invention, i.e., why Applicants believes the claimed invention is useful. It is important to note that there is *no* requirement under 35 USC 101 for the utility to be expressly set forth in the claims. For example, compound claims

need only recite the structure of the compound. A proper 101 analysis includes a review of the specification to see if there is an asserted utility, and if so, the requirements of 35 USC 101 have been met.

Second, Applicants respectfully submit that a complete disclosure has been set forth in the specification as required by this section of the MPEP. Specifically, the present inventors have described on pages 1 and 2 of the specification that the inventive computer program is intended to be capable of canonically coding stereochemical information for predicting NMR chemical shifts. Also, the inventive program provides a means for notating the stereostructural environment into computer readable data.

MPEP 2106(II)(A) further goes on to state:

> The claimed invention as a whole must produce a "useful, concrete and tangible" result to have a practical application.

What exactly constitutes a "useful, concrete and tangible" result is taken on a case by case basis. However, Applicants respectfully submit that certain cases providing guidance would lead the Examiner to the conclusion that the claimed invention produces such a result. Specifically, *State Street Bank & Trust Co. v. Signature Financial Group Inc.*, 47 USPQ2d 1596 (CAFC 1998) has facts similar

to the present case. In this case, the court applied a test for compliance with Section 101 which was originally described in *In re Freeman*, 573 F.2d 1237, 197 USPQ 464 (CCPA 1978) but was modified by *In re Walter*, 618 F.2d 758, 205 USPQ 397 (CCPA 1980). The test has been thus articulated:

> First, the claim is analyzed to determine whether a mathematical algorithm is directly or indirectly recited. Next, if a mathematical algorithm is found, the claim as a whole is further analyzed to determine whether the algorithm is "applied in any manner to physical elements or process steps," and, if it is, it "passes muster under Section 101."

Using this test, the court found that the transformation of data, representing discrete dollar amounts, through a series of mathematical calculations into a final share price, "constitutes a practical application of a mathematical algorithm, formula, or calculation, because it produces 'a useful, concrete and tangible result' -- a final share price momentarily fixed for recording and reporting purposes...". *State Street*, 47 USPQ2d at 1601. *State Street* holds that even if the useful result is expressed in numbers printed or recorded on some computer readable recording media, such as price, profit, percentage, cost, or loss, it is statutory subject matter.

In the present case, the useful result is numbers/symbols which are printed or recorded on computer readable recording media and therefore the claims recite statutory subject matter.

With regard to the rejection under 35 U.S.C. 112, 1st paragraph, the Examiner only provides a brief description as to the basis for the rejection. Specifically, the Examiner indicates that since the presently claimed invention does not have a specific utility, then it naturally follows that the present invention lacks an enabling disclosure as required under 35 U.S.C. 112, 1st paragraph.

Applicants respectfully submit that for reasons set forth above, the present invention satisfies the requirements of 35 U.S.C. 101.

On this matter, MPEP 2106(V)(B)(2) states that the

> "specification should disclose how to configure a computer to possess the requisite functionality or how to integrate the programmed computer with other elements of the invention, unless a skilled artisan would know how to do so without such disclosure." (Emphasis in original).

In the present case, the Examiner has given no reasons why the skilled artisan would not know how to configure a computer to possess the requisite functionality in light of the inventive disclosure.

As stated by the Federal Circuit:

> When rejecting a claim under the enablement requirement of Section 112, the [Patent Office] bears an initial burden of setting forth a reasonable explanation as to why it believes that the scope of protection provided by the claim is not adequately enabled by the description of the invention provided in the specification of the

> application; this includes, of course, providing sufficient reasons for doubting any assertions in the specification as to the scope of enablement. *In re Wright*, 999 F.2d 1557, 27 USPQ 2d 1510, 1513 (Fed. Cir. 1993).

Applicants respectfully submit that the Examiner has made unsubstantiated assertions which lack evidentiary support. The Examiner has not established a *prima facie* case of nonenablement, i.e., the Examiner has not provided sufficient reasons why Applicants' assertion is incorrect that the present program is useful to prepare stereostructural information regarding atoms in a molecule which is utilizable as computer readable data for applications such as calculating NMR chemical shifts.

In addition, the skilled artisan is aided by the instant disclosure and the advanced state of the art to perform the instant invention. As described in the present specification beginning at page 9, line 18, the skilled artisan is shown specific non-limiting examples of using the inventive program to obtain stereostructural information of certain sugars.

Thus, Applicants respectfully submit that the test for enablement has been met and withdrawal of the rejection under 35 USC 112, first paragraph is respectfully requested.

Issues under 35 U.S.C. 112, 2nd paragraph

Claims 1-12 are rejected under 35 U.S.C. 112, 2nd paragraph as being indefinite.

Applicants respectfully traverse the rejection.

Applicants herein incorporate by reference the comments made regarding this issue in the January 7, 2004 Reply in their entirety.

Conclusion

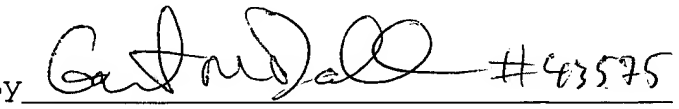
In view of the above amendments and comments, Applicants respectfully submit that the claims are in condition for allowance. A notice to such effect is earnestly solicited.

Should there be any outstanding matters that need to be resolved in the present application, the Examiner is respectfully requested to contact **Garth M. Dahlen, Ph.D., Esq.** (Reg. No. 43,575) at the telephone number of the undersigned below.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37 C.F.R. §§ 1.16 or 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By  #43575
Marc S. Weiner, #32,181

Garth M. Dahlen, #43,575

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

MSW/GMD/bmp
0051-0155P

Attachments: (1) Abe et al., J. Chem. Inf. Comput. Sci., Vol. 24, No. 4, pg 212, 1984

Reprinted from Journal of Chemical Information and Computer Sciences, 1984, 24, 212

A Convenient Notation System for Organic Structure on the Basis of Connectivity Stack

HIDETSUGU ABE, YOSHIHIRO KUDO,[†] TOHRU YAMASAKI,[‡] KAZUO TANAKA,[§]
MASAHIRO SASAKI, and SHIN-ICHI SASAKI*

Laboratory for Chemical Information Science, Toyohashi University of Technology, Toyohashi, Aichi, Japan 440

Received August 8, 1983

A convenient notation system for organic structures has been developed for the application of the connectivity stack. A notation arbitrarily encoded for a structure by a user through a rather simple procedure using 35 codes, which have been previously prepared, is automatically canonicalized in a computer. The notation given by the user is standardized according to the rules for rearranging the codes into a dictionary order. The connectivity stack is estimated for each of the standard notations and its permuted derivatives. The notation whose stack is the largest amount is decided to be canonical. This notation method will be widely applicable in the field of structure manipulation because of its extreme simplicity.

Several methods for the representation of organic structures have been investigated for computer-aided storage and retrieval of the structures.[1] Linear notations and connection table methods are two major techniques for the topologically unambiguous and unique representation of chemical structures.[2] The connection table descriptions specify all the atoms of a molecule (hydrogen is often suppressed) and may explicitly describe the connectivity of each atom. On the other hand, one of the features of the linear notation method is that chemical structure can be expressed more compactly by the use of letters, numerals, and some symbols. The number of letters, numerals, and symbols used to represent a structure is, in general, much fewer than the number of atoms included in the structure. According to such compactness, the linear notation method seems to be more preferable than the connection table method for compilation of a vast number of structures to be treated in a computer. However, the procedure for canonicalization of a linear notation is generally so tedious and complicated that users hesitate to adopt the methods.

In this paper, we present a new notation system on the basis of a "connectivity stack", which has been published by Y.K. and S.S.[3] The notation system, CANOST (autoCANOnicalization system for organic STructures), has the following features. (1) The notation given arbitrarily by the user through rather simple procedures described later is automatically canonicalized in a computer. (2) Thirty-five symbols expressing atoms, atomic groups, ionic charges, and others as listed in Table I are used to make the arbitrary notation. Two or three hours is normally sufficient to learn how to encode chemical structures for even a beginner in chemistry. (3) Though most of structures are expressed with the 35 items, any other symbols consisting of up to four letters may be added if necessary. (4) The notation can be easily converted into a corresponding connection table. The latter, in some cases, is more usable and convenient than the linear notation for computer-aided manipulation of structures.

Table I. Code of Substructure in CANOST[a]

no.	substructure	code	no.	substructure	code
1	—C≡	T	16	—C—O	VD
2	HC≡	T1	17	—O—	Q
3	=C=	DD	18	—OH	Q1
4	>C=	D5	19	=O	QD
			20	—F	LF
5	—CH=	D1	21	—Cl	LC
6	H_2C=	D2	22	—Br	LB
7		C	23	—I	LJ
			24[e]	single bond	SG
			25	cation	+
8	>C<	C1	26	anion	−
			27	radical	·
9	$—CH_2—$	C2	28	chelation	/
10	$—CH_3$	C3	29	other atom	X
11[b]		Y	30[f]		XR
12[b]		Y1	31		XW
13[c]		YT	32	=X	XD
			33	—X=	XX
			34	≡X	XT
14	>C=O	V	35	XH_p	XP
15	—CHO	V1			

[a] Aromatic carbon without hydrogen. [b] Aromatic carbon with hydrogen. [c] –C(OH)–C(O)– in troponoid. [d] Prepared for connecting D1 to clearly express conjugated double bond (see Figure 3). [e] Non-carbon atom in aromatic structure.

GENERAL ENCODING PROCEDURES

The following describes how to encode a chemical structure into CANOST notation.

Step 1. Select proper symbols from Table I for the atoms and atomic groups in a structure concerned. If two or more alternative encodings are possible for the structure, the one

[†] Present address: Faculty of Engineering, Yamagata University, Yonezawa, Yamagata, Japan 992.
[‡] Present address: Mitsui Petrochemical Industry Ltd. Co., Iwakuni, Yamaguchi, Japan 740.
[§] Present address: Asahi Research Center Co. Ltd., Uchisaiwai-cho, Chiyoda, Tokyo, Japan 100.

0095-2338/84/1624-0212$01.50/0

NOTATION SYSTEM FOR ORGANIC STRUCTURE

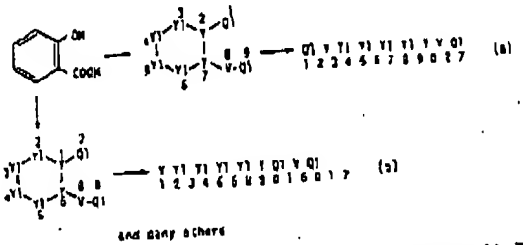


Figure 1. Arbitrary and noncanonical notation of salicyclic acid. Bar and hyphen are used only for graphical representation for user's convenience and should not be confused with anion (−) symbol in Table I or Figure 4. This applies corresponding to the graphical notation in Figures 2 and 3.

consisting of the smallest number of symbols should be selected.

Step 2. Numerals (1, 2, 3, ...) are allocated consecutively to every symbol. Since there is no particular rule for the numbering, it is allowed to start at any symbol.

Step 3. Place all symbols in a line according to the order of the numbers assigned above.

Step 4. To express connections of symbols, arrange the corresponding numerals in a line. There are two rules for arranging the numerals; one is that continuous arrangement of the numerals indicates the connection of the corresponding symbols, and the other is that zero is used when the connection is suspended.

Figure 1 illustrates the process of arbitrary and noncanonical notation of a structure with salicyclic acid as an example. In case (a) in the figure, first, six aromatic carbons, two hydroxyls, and one carbonyl of the structure are replaced by the corresponding symbols, four Y1's, two Y's, two Q1's, and one V, respectively, selected from the symbols listed in Table I. Then, numerals (1, 2, 3, ..., 9) are assigned arbitrarily to all the symbols. Second, the symbols are placed in a line, from left to right. To express connection of the symbols, the numerals are placed as shown in the figure. The connection of Q1, Y, Y1, Y1, Y1, Y1, Y, V, and Q1 is indicated by continuous arrangement of numerals, 1–9. After 9, 0 (zero) comes to indicate that the connection is suspended here. Then, again, 2 and 7 are placed to show two Y's numbered 2 and 7 are connected to each other. Another arbitrary notation (b) is also given in Figure 1, and many other notations, are, of course, possible for the structure, according to arbitrariness of an individual user. Any notation can be transformed into a canonical form along the algorithm mentioned later.

ADDENDUM RULES FOR STEP 1

Although these procedures for noncanonical notation of salicyclic acid are easily applied to all other structures, it will be necessary to add some more general rules in addition to the above-mentioned rules for the notation.

(1) Symbols for the atoms and atomic groups containing hetero elements other than oxygen and halogens are expressed by symbols 29–35 in Table I. Users can make the proper symbols as they like by substituting X with actual elemental symbols N, S, P, or other hetero elements. For example, nitro group and amino group are represented as NW and N2, respectively, by substituting X in symbols 31 and 35 with N. Another examples are shown in Figure 2.

(2) When a structure contains conjugated double bonds, the structure is expressed in either way with or without SG (24) as shown in Figure 3. SG is prepared to help a user out of his confusion of the encoding, especially, in the case of highly conjugated structure, e.g., carotenoid. Since this is a dummy, this is neglected in the process of canonicalization and does not appear in the final canonical notation.

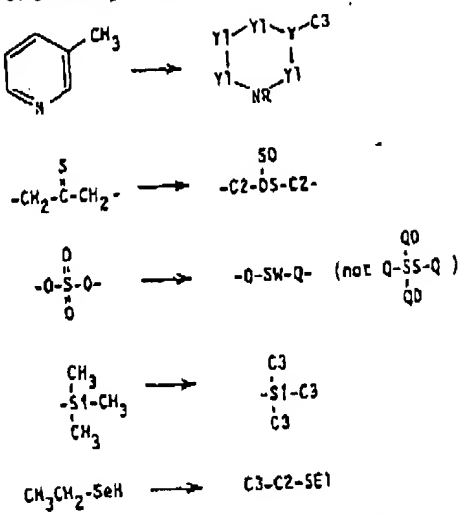


Figure 2. Encoding of structure with heteroatom.

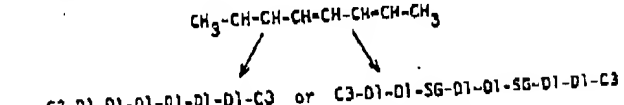


Figure 3. Encoding of structure with conjugation.

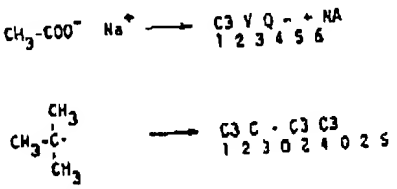


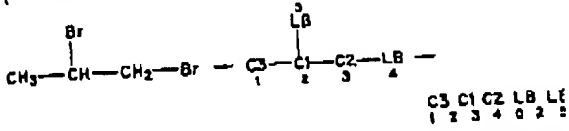
Figure 4. CANOST notation of salt and radical.

(3) Salt and radical are expressed as shown in Figure 4.

ALGORITHM OF CANONICALIZATION

The formal algorithm for the canonicalization has already been published by Y.K. and S.S.[3b] Thus, the routes to reach canonical form from noncanonical form are described with the example 1,2-dibromopropane. Before that, it is still necessary to mention two important principles for the canonicalization. The first is that CANOST codes consisting of a noncanonical notation should be rearranged in the order of Z Y X ... N M L ... C B A 9 8 ... 3 2 1 0 / ● − + □ (□ stands for a space). The second is that the notation whose connectivity stack has the largest binary value among the possible notations for a certain structure expressed by CANOST codes is defined to be canonical form.

Now, let us consider the case of 1,2-dibromopropane, as a simpler example. One of the noncanonical notations of the structure might be expressed by a user by the procedure previously mentioned as



The above notation is rearranged according to the first principle, and simultaneously, the code number is modified by reallocation of numerals 1–5; the result is

1,2-Dibromopropane is expressed by two possible connectivitie

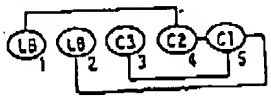
(1)

connectivity matrix	LB	LB	C3	C2	C1
LB	x	0	0	0	1
LB		x	0	1	0
C3			x	0	1
C2				x	1
C1					x

connectivity stack .0 0 0 0 1 0 1 0 1 1

CANOST connectivity LB LB C3 C2 C1 1 5 3 0 2 4

(2)



connectivity matrix	LB	LB	C3	C2	C1
LB	x	0	0	1	0
LB		x	0	0	1
C3			x	0	1
C2				x	1
C1					x

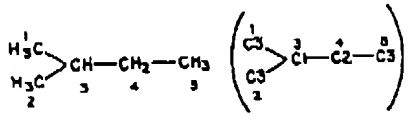
Connectivity stack .0 0 0 1 0 0 0 1 1 1

CANOST connectivity LB LB C3 C2 C1 1 4 5 2 0 3 5

Figure 5. Two potential connectivities of CANOST codes of 2,3-dibromobutane.

of these codes as shown with (1) and (2) in Figure 5. According to the second principle, it may be decided whether (1) or (2) is canonical by comparing their connectivity stacks. As connectivity stack is defined to be a numeral string derived by arranging the off-diagonal upper triangle of a connectivity matrix in row by row manner, .0000101011 and .0001000111 are obtained for the connectivity matrixes of (1) and (2), respectively, as shown in Figure 5. Thus, the latter (2) is decided for the canonical form of the structure, which is expressed as LB LB C3 C2 C1 1 4 5 2 0 3 5. The rule to make a line of numerals, 1 4 5 2 0 3 5, indicating connectivity of codes of (2) is that LB (1) connecting to C2 (4) makes 1 4 and C2 (4) connecting to C1 (5) makes 4 5. C1 (5) connects to LB (2) and C3 (3). In such a case, the smaller numeral, 2, comes after 5. So far, 1 4 5 2 is made, and here, the linear connection is suspended by placing a zero (0) next. C3 (3) and C1 (5) makes 3 5, as the smaller numeral is always first. Thus, 1 4 5 2 0 3 5 is made to express the connection of all the codes, LB, LB, C3, C2, and C1. To have the canonicalization processed in a computer, the codes used to express structure are converted into eight-digit numerals by refering to Table II.

Let us show the processes of canonicalization performed by a computer. The first example is 2-methylbutane. Noncanonical notation, C3 C3 C1 C2 C3 1 3 4 5 0 2 3, originated from



is sent to a computer, where the following things are carried out.

Table II. Numerals Given for Elements of CANOST Codes

symbol	numeral	symbol	numeral	symbol	numeral
□	32	9	57	N	78
+	43	A	65	O	79
-	45	B	66	P	80
.	46	C	67	Q	81
/	47	D	68	R	82
0	48	E	69	S	83
1	49	F	70	T	84
2	50	G	71	U	85
3	51	H	72	V	86
4	52	I	73	W	87
5	53	J	74	X	88
6	54	K	75	Y	89
7	55	L	76	Z	90
8	56	M	77		

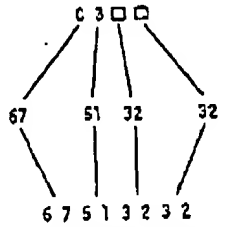


Figure 6. Numeralization of code C3. Two spaces always follow after code, which are replaced by 3 2 3 2.

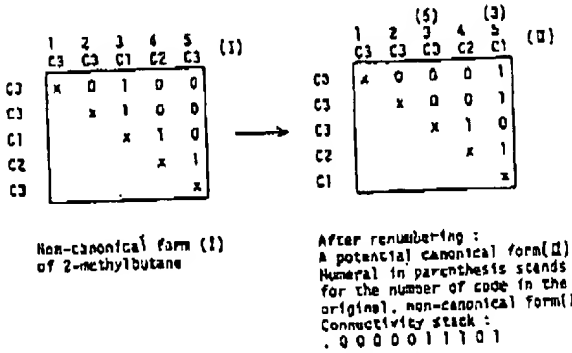
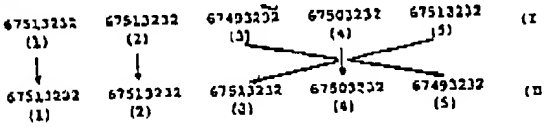


Figure 7. Generation of a possible canonical form by renumbering of code number.

(1) Each code of the input noncanonical notation inputted is converted into the corresponding numeral by applying Table II. For instance, C3 is replaced by 6 7 5 1 3 2 3 2 as shown in Figure 6. Thus, the noncanonical notation of the structure is finally represented like 67513232 67513232 67493232 67503232 67513232 1345023. These numeralized codes are rearranged in descending order in accordance with their amounts:



By this renumbering, the original connectivity matrix based on the connection of I is converted into a new connectivity matrix based on II, which corresponds to one of the potential canonical forms (Figure 7).

(2) The potential canonical form provided by the renumbering is one of the candidate canonical notations and its connectivity stack is .0000011101 as shown in Figure 7. If there are more than one particular CANOST code present in a structure, exhaustive permutation of their order should be done to look for the connectivity matrix with the large

Table III. Six Possible Permutations of Three C3's That Are Labeled with 1, 2, and 3

	II	IV	V	VI	III	VII
C3	1	1	2	2	3	3
C3	2	3	1	3	1	2
C3	3	2	3	1	2	1
C2	4	4	4	4	4	4

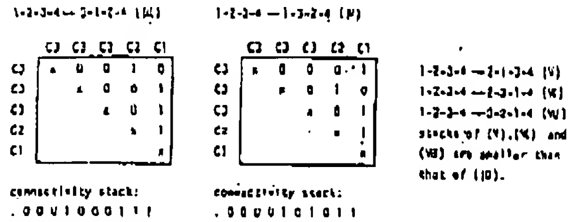


Figure 8. Comparison of connectivity stacks resulted by permutation of three C3's.

	V	C2	C1	C2	C1	C2	C3	C3 (VIII)
V	x	1	0	0	1	0	0	0
C2		x	1	0	0	0	0	0
C1			x	1	0	0	1	0
C2				x	1	0	0	0
C1					x	1	0	1
C2						x	0	0
C3							x	0
C3								x

Figure 9. Connectivity matrix of a noncanonical notation of 3,5-dimethylcyclohexanone.

connectivity stack. In this example, three C3's are present; therefore, six operations, 1 2 3 4 (II), 1 2 3 4 → 1 3 2 4 (IV), 1 2 3 4 → 2 1 3 4 (V), 1 2 3 4 → 2 3 1 4 (VI), 1 2 3 4 → 3 1 2 4 (III), and 1 2 3 4 → 3 2 1 4 (VII), are necessary for the complete permutation (Table III).

(3) The permuted results are shown in Figure 8. When the numbers of codes in II (1 2 3 4) are rearranged into 3 1 2 4 (III), the connectivity stack of III is found to be the largest among those of II (Figure 4), IV, V, VI, and VII. Thus, III is decided to be canonical for 2-methylbutane.

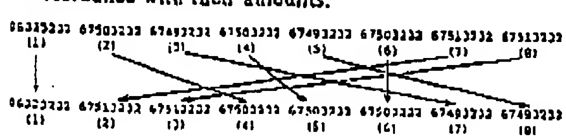
(4) Matrix III is then converted into canonical CANOST connectivity according to the procedure shown in Chart I.

(5) The output is the canonical CANOST notation of 2-methylbutane, C3 C3 C3 C2 C1 1 4 5 2 0 3 5. Example 2 is 3,5-dimethylcyclohexanone. The noncanonical notation given arbitrarily for this structure, V C2 C1 C2 C1 C2 C3 C3 1 2 3 4 5 6 1 0 3 7 0 5 8, is sent to the computer.

(6) All the codes are replaced by eight-digit numerals by referring to Table II: V C2 C1 C2 C1 C2 C3 C3, 86323232 67503232 67493232 67503232 67493232 67503232 67513232 67513232.

(7) The above noncanonical notation given arbitrarily is transformed into the corresponding connectivity matrix (VIII) (Figure 9).

(8) Numeralized codes are rearranged in descending order in accordance with their amounts:



Here, renumbering of codes, steps 2–8, is carried out.

(9) By the above renumbering, the original matrix (V) generated at step 7 is transformed into a new matrix (IX) in Figure 10 that becomes a standard to look for the canonical notation with the largest connectivity stack.

(10) As similarly in example 1, the same types of codes (two C3's, three C2's, and two C1's in the example) are exhaustively permuted. The computer found that the largest connectivity stack is obtained when the connection of 1 2 3 4 5 6 7 8 of the standard (IX) is rearranged like 1 2 3 4 6 5 7 8 (X) (Figure 10).

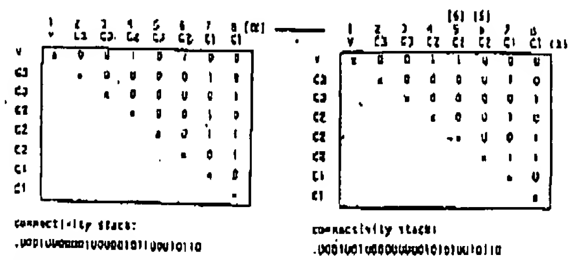
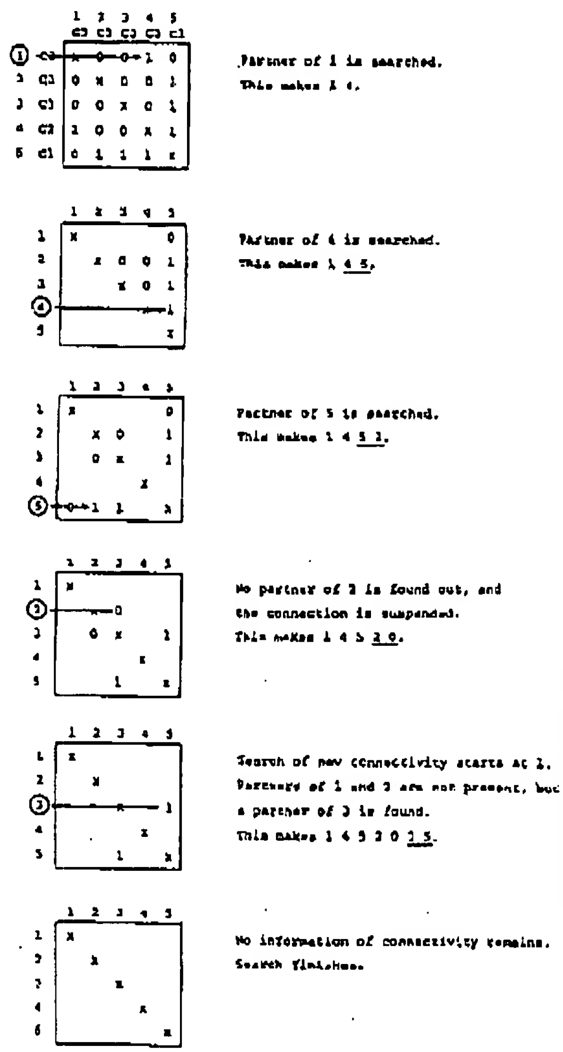


Figure 10. Comparison of the connectivity stacks resulted by permutation of identical nodes. Only two results are shown. Stack of X is the largest among those of 24 connectivity matrixes. Numerals in parentheses indicate the number of code in IX.

Chart I



6/7

(11) V C3 C3 C2 C2 C2 C1 C1 1 4 7 2 0 1 5 8 3 0 6 7 0 6 8 is uniquely presented as the canonical notation of 3,5-dimethylcyclohexanone.

APPLICATION OF THE CANOST NOTATION SYSTEM

The present notation system is an undoubtedly convenient tool for a variety of data bases in which manipulation of structural formulas is required. Actually, the system has been applied to the system SPIRES (SPectral Information REtrieval System), in which a ^{13}C NMR data base system is contained. The data base system, an interactive retrieval system of structure (substructure)–spectral information, has already been reported briefly,[4] and more detail will be presented in the following paper.[5]

A function to represent stereochemical structure has not yet been included with the system. The problem, however, will be solved without much difficulty by adding new symbols indicating stereochemistry such as, for example, E and Z for geometrical and R and S for configurational isomers and by slightly modifying the program.

REFERENCES AND NOTES

(1) Zupan, J.; Heller, S. R.; Milne, G. W. A.; Miller, J. A. *Anal. Chim. Acta* 1978, *103*, 141. Bremser, W. *Anal. Chim. Acta* 1978, *103*, 355. Wipke, W. T.; Heller, S. R.; Feldmann, R. J.; Hyde, E., Ed. In "Computer Representation and Manipulation of Chemical Information"; Wiley-Interscience: New York, 1974.

(2) Lynch, M. F.; Harrison, J. M.; Town, W. G.; Ash, J. E. "Computer Handling of Chemical Structure Information"; Macdonald: London, 1971. Davis, C. H.; Rush, J. E. "Information Retrieval and Documentation in Chemistry"; Greenwood Press: Westport, CT, 1974.

(3) (a) Kudo, Y.; Yamasaki, T.; Sasaki, S. *J. Chem. Doc.* 1973, *13*, 225. (b) Kudo, Y.; Sasaki, S. *J. Chem. Doc.* 1974, *14*, 200. (c) Kudo, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* 1976, *16*, 43. (d) Sasaki, S.; Abe, H.; Hirota, Y.; Ishida, Y.; Kudo, Y.; Ochiai, S.; Saito, K.; Yamasaki, T. *J. Chem. Inf. Comput. Sci.* 1978, *18*, 211.

(4) Abe, H.; Sasaki, S.; Tanaka, K.; Osada, H. *CODATA Bull.* 1981, *No. 40*, 31.

(5) Abe, H.; Hayasaka, H.; Miyashita, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* 1984, *24*, 216–219.

7/7